Gradient perception of intonation

Caroline Féry, Ruben Stoel

University of Potsdam

Abstract[1]

Phonologists usually reserve the term gradience for the distinction between phonetics and phonology. Phonetics is gradient and phonology is categorical. But in recent years, phonotactic patterns have been found to be gradient both along frequency counts and grammaticality judgments (see for instance Pierrehumbert 2003, Frisch & Stearns, this volume, and Jurafsky 2003). This chapter addresses the question of gradience in a different area of phonology, namely the tonal patterns of utterances. It concentrates on the graded acceptability of tonal patterns in contexts, and experimentally shows that acceptability judgments about tonal contours in declarative sentences are gradient in an interesting way. A sentence's most 'neutral' tonal patterns (Höhle 1982) is the one expressed in an 'all-new' or 'broad-focus' context, the "Unmarked Prosodic Structure" (UPS) of the sentence. In German, it is also the pattern found in main clauses with a topicalized subject and a focused VP. Marked prosodic patterns, like those uttered in a narrow focus context, require a special information structure, and thus have a restricted context of use. As a result, tonal contours are not gradient as such, but only in association with texts and information structure. In the experiment reported in the paper, we study the pattern of acceptability of neutral and marked intonation patterns in context. Though the research reported in this chapter is limited to declarative sentences, we are confident that other kinds of tonal patterns are gradient in a similar way.

1 Introduction

Many phonologists associate the term 'gradience' with the distinction between phonology – which is supposed to be categorical– and phonetics –which is supposed to be gradient– (see Cohn, this volume, for a review of the issues associated with this distinction). In recent years, a different role for gradience in phonology has emerged: the well-formedness of phonological structures has been found to be highly gradient in a way that correlates with their frequency. In their chapter, Frisch & Stearns (this volume) show that phonotactic patterns, like consonant clusters and other segment sequences, as well as morphophonology, word-likeness, etc. are gradient in this way. The examination of large corpora is a reliable indicator of relative frequency. Crucially, the less frequent sequences are felt by speakers to be less prototypical exemplars of their category. In grammaticality judgment tasks, word-likeness tasks, assessment of novel words etc., less frequent items are likely to get lower grades than more frequent ones. In short, speakers reproduce in their judgments the pattern of relative frequency that they encounter in their linguistic environment. In light of this well-documented (see Frisch & Stearns, this volume and references cited there), but controversial result, the question has arisen for some phonologists as to the need of a grammar operating with abstract phonological categories, like features and phonemes. In their opinion, if phonotactic distribution is learnable by executing probabilistic generalizations over corpora, the only knowledge we need in order to elaborate 'grammars' may turn out to be a stochastic one. But before we can take a stand on this important issue in a competent way, we need to be well-informed on other aspects of the phonology as well.

In this chapter, we take a first step and investigate the question of whether intonational contours are gradient in the same way that segment sequences are. Is it the case that more frequent tonal patterns are more acceptable than less frequent ones? We use the term gradience in the sense of gradient acceptability.

Unfortunately, for a number of reasons, large corpora – at least in their present state – are useless for the study of tonal patterns frequencies. One of the reasons relates to the analysis and annotation of tonal patterns. Scholars not only disagree on the kinds of categories entering intonation studies but also on the annotation for 'rising contour', 'fall-rise' etc. Melodies – like Gussenhoven's (1984, 2004) nuclear contours or the British school's 'heads' and 'nuclei' – may well exist as independent linguistic elements, but they are not transcribed uniformly. Even though autosegmental-metrical representations of tonal contours, like ToBI (Beckman & Ayers 1993, Jun 2005) are evolving to become a standard in intonation studies, they are not sufficiently represented in corpora. Most large corpora consist of written material anyway, and those which contain spoken material generally only display segmental transcription rather than tonal.

In short, the development of corpora which are annotated in a conventional way for intonation patterns is an aim for the future, but as of now, it is simply not available for German.

As a result, we must rely on the intuition of speakers. The questions we address in this chapter are: Which tonal contours are accepted most? Which are less accepted? We will see that the question must be made precise in the following way: given a certain syntactic structure, is there a contour which is accepted in the largest set of contexts? And this is related to the question of pitch accent location. Which constituents are expected to be accented? Which accent structure is the least marked, in the sense of being accepted in in the most contexts? Are some accent patterns (tonal patterns) 'unmarked' (more frequent, acquired earlier, but also accepted more easily) in the same sense as consonant clusters or other segment sequences are?

Below, we present the results of a perception study bearing on tonal contours. But before we turn to the experiment, we first sum up some relevant issues in the research on prosody and situate our research in this broader context.

2 Prosody and intonation

Prosody plays a crucial role in communication. To begin with, we partition our utterances in prosodic chunks, like phonological phrases and intonation phrases, which correspond to syntactic constituents (Nespor & Vogel 1986, Truckenbrodt 1999) or information structural blocks (Vallduví 1992). These phrases, which help both speakers and hearers structure the discourse, are signaled phonetically by boundary tones, segmental lengthening or some other phonological cues. A second factor playing a role in phonological patterning is the distribution and form of pitch accents, associated with prominent syllables. A syllable may be prominent if it is the bearer of the lexical stress of a word or of a larger constituent which is itself prominent. A speaker may decide to speak about some object in her surroundings or an object she knows about, and decide to focus on one property of this object. Or she may answer a question asked by a protagonist because she feels she has to deliver some bit of information. In other words, prominence may be assigned to some linguistic constituents because of contextual or cognitive reasons (Bolinger 1972). The other reason to assign a pitch accent to a syllable is purely grammatical. An internal argument of a German predicate + argument complex, for example, may receive a pitch accent, and the verb may be unaccented. Still, the whole phrase may be prominent (see Bierwisch 1968, Schmerling 1976, Gussenhoven 1983, 1992, von Stechow & Uhmann 1986, Cinque 1993, Féry & Samek-Lodovici 2006, among others).

In Standard German, nuclear accents (the final or most prominent accents of an intonation phrase) are either bitonally falling, HL, or rising, LH, whereas prenuclear accents can be rising or falling as well or monotonally high (H) or low (L) (see Féry 1993, Grabe 1998, Grice, Baumann and Benzmüller 2003, Peters 2005 for phonological studies of intonation of standard German). Prosodic phrases may be terminated with a boundary tone, which is written with a subscripted $_P$ for a phonological phrase, and a subscripted $_I$ for an

intonation phrase (following Hayes & Lahiri's 1991 notation). For the sake of illustration, two pitch tracks of a sentence used in the experiments described below are shown with their phonological tone structure. The first pitch track, in figure 7.1, is equivalent to a wide-focused realization with two pitch accents, a rising one on the subject *Ruderer*, and a falling one on the object *Boote*. The verb, adverb and particle *mit* are unstressed. This realization may be dubbed 'unmarked prosodic structure' (UPS, see below). It is expected to be the most frequent one, and thus, the most widely accepted pattern for such a declarative sentence. In German, a topic-focus realization, in which the subject is topicalized and the remainder of the sentence is focused, is identical to a wide-focused realization.

INSERT FIGURE 7.1 HERE

The second pitch track (figure 7.2) shows a marked pattern, with just one pitch accent located early in the sentence. This kind of pattern is expected to be confined to special contexts, in particular those eliciting a narrow focus on the subject.

INSERT FIGURE 7.2 HERE

It is not possible to investigate the gradience of tonal patterns out of context. Tonal patterns do not exist as pure melodies: they need to be interpreted as linguistic units, thus as pitch accents or as boundary tones. This can only happen when tonal excursions are associated with text. Moreover, tonal contours are more or less marked only when they are

associated with specific locations in a sentence, since accent locations are dependent on syntax and information structure. We introduce 'focus projection' briefly in the next section, but have no space to develop all arguments for this phenomenon (see Selkirk 1995, Schwarzschild 1999 and Féry & Samek-Lodovici 2006 among others). We propose the concept of 'Unmarked Prosodic Structure' (UPS, Féry 2005) as the intonation used when the sentence is realized in a whole-focused environment. It refers to the phrasing and the tonal contour projected when the speakers have no clue about the context. Unmarked Prosodic Structure relies solely on the syntactic structure. A tonal contour compatible with unmarked prosody is expected to be acceptable in more environments than other, more marked contours.

3. Previous studies on gradient tone perception

Few studies, if any, have explicitly addressed the gradience of intonational contours, so we cannot base our work on a rich empirical basis. There are, however, quite a number of studies investigating the question of categories in intonational morphemes, which have found more or less gradient accents or boundaries.[2] The most relevant studies for our aim have looked at the adequacy of pitch accent patterns in some specific contexts.

The issue of the location of pitch accents and their role for the focus structure has been investigated for English by Gussenhoven (1983) and Birch & Clifton (1995), among others, who examine the role of prenuclear accents on the verb in a VP consisting of a verb plus an argument (or an adjunct by Gussenhoven) in English. Gussenhoven's (1983) sentence accent assignment rules (SAAR) predict that in a focused predicate argument complex, only the argument needs to be stressed, but that a prenuclear accent can be added freely on a verb without impairing processing. In a verbal phrase, by contrast, both the verb and the adjunct need to be stressed. Gussenhoven himself finds confirmation of this prediction in experimental work. In mini-dialogues such as (1), there is a difference between the focus structure of (1a) and (1b). In (1a), the whole VP *share a flat* is focused, whereas in (1b) only

the direct object is focused, the difference being elicited by the preceding question. The same kind of contrast is obtained in the dialogues in (2) which contain a verb followed by an adjunct.

(1) Verb and argument

a. C: Do you live by yourself?

b. C: I hate sharing things, don't you?

c. U: I share a flat. (the whole VP or the argument NP is focused)

(2) Verb and adjunct

a. C: Where will you be in January?

b. C: Where will you be skiing?

c. U: We will be skiing in Scotland. (the whole VP or the adjunct PP is focused)

Gussenhoven cross-spliced questions and answers, spoken by native speakers, so as to obtain both answers in both contexts. Subjects then had the task of deciding which of the two answers was the more appropriate response to the preceding question. Gussenhoven found that the presence of an accent on the verb in addition to the expected accent on the object in (1) does not change the acceptability of the pitch accent structure, and that this held in both narrow and broad focused contexts. The speakers did not do better than by chance when required to choose between the two contexts on the basis of such an accent pattern. But in (2), the absence of a stress on the verb in (2a) was an indicator that the verb had to be given (and thus not focused), so that the speakers did better than in the predicate-argument condition in the same task. The reliability of the accent on the verb in deciding for the wide-focus context depended gradiently on the number of unstressed syllables intervening between the two accents.

Birch & Clifton (1995) conducted similar experiments, but obtained slightly different results. They also prepared matched and mismatched pairs of questions and answers. An example of a dialogue set is reproduced in (3). Only the pairs QA/R1 and QB/R3 are perfectly matching, all others are predicted to be more or less deviant along the same lines as those just explained, though the authors acknowledge that QA/R2 could be as good as QA/R1 if SAAR make the right predictions.

(3) a. Questions QA: Isn't Kerry pretty smart?

QB: Isn't Kerry good at math?

b. Responses R1: Yes, she TEACHES MATH.

R2: Yes, she teaches MATH.

R3: Yes, she TEACHES math.

In judgment and decision tasks, Birch & Clifton found that as an answer to question QA, speakers prefer R1, with two accents, over R2, with just one accent on the argument NP. The difference was small but significant. And unsurprisingly, R3 was by far the preferred answer to QB. All other pairs obtained poorer scores. In a second experiment, speakers had to decide how well the pairs made sense. In this case, the results for QA were similar to those of Gussenhoven: There was no difference between a sentence with two accents (R1) and a sentence with just one accent on the argument (R2).[3]

These results, as well as other perception experiments bearing on the location of pitch accents conducted for Dutch (Nooteboom & Kruyt 1987, Krahmer & Swerts 2001) and for German (Hruska et al. 2001) show that, for these three languages at least, a prenuclear accent is readily acceptable, but that a postnuclear one is less easily accepted and that accents on narrowly focused items in an otherwise non-nuclear position are more readily perceived than accents on words accented per default in their unmarked accent pattern.

Nooteboom & Kruyt (1987) rightly explain the acceptability of a prenuclear accent in terms of topicalizing or thematicizing the bearer of such an accent, and observe that a sentence with a supplementary prenuclear accent can get an interpretation in which the prenuclear accent is information structurally prominent.

In psycholinguistic experiments studying the role of prosody in disambiguating syntactic structures (see for instance Lehiste 1973, Kjelgaard & Sheer 1999 and Schafer et al. 2000), garden path sentences or sentences with an ambiguous late or early closure/attachment have been tested. These experiments deliver gradient results correlating with the strength and the location of boundaries. Comparing the two realizations of the sentences in (4), there is no doubt that intonation can disambiguate the readings. (4a) is realized as one Intonation Phrase, but in (4b), an Intonation Phrase boundary is located after *heiratet*, which is then understood as an intransitive verb. Much more subtle is the question of whether prosody can help with the sentence in (5). In one reading, it is the woman who lives in Georgia, and in the other reading, her daughter. The phrasing, in the form of a Phonological Phrase boundary, is roughly the same in both readings. Nevertheless, it is possible to vary the quantity and the excursion of the boundary tone in such a way that the preference for one or the other reading is favored.

L*H L*H H*L L_I

(4) a. [MaRIa heiratet MARtin NICHT]$_I$

'Mary does not marry Martin'

L*H H*L L_I L*H H*L L_I

b. [MaRIa HEIratet]$_I$ [MARtin NICHT]$_I$

'Mary gets married. Martin does not.'

(5) [[Ich treffe mich heute]$_P$ [mit der Tochter der Frau]$_P$]$_I$ [[die in Georgien lebt]$_P$]$_I$

'I am meeting today with the daughter of the woman who lives in Georgia.'

We are only marginally interested in syntactic disambiguation in this chapter. Rather, our experiment aimed at testing the gradience of German intonational structures. Our experiment differs from the ones conducted by Gussenhoven and Clifton & Birch in a crucial way: several parameters were systematically varied: sentence type, context and tonal contours. We were explicitly interested in finding out whether some kinds of intonation patterns are more acceptable than others and whether gradience can be observed in the domain of tonal contours.

4 Experiment

4.1 Background

The experiment reported in this section was intended to elucidate the question formulated above: How gradient are tonal contours? We would like to understand what triggers broad acceptance for intonational patterns. To this aim, we used three different kinds of sentences, which were inserted in different discourse contexts, and cross-spliced. If an effect was to be found, we expected it to be of the following kind: We expected the unmarked tonal contours to be generally better tolerated than the marked ones.

The hypothesis can be formulated as in (6).

(6) Unmarked Prosodic Structure (UPS) Hypothesis

An unmarked prosodic structure, i.e. a prosodic structure adequate in a broad focus environment, is readily accepted. It can be inserted successfully in more environments than a marked prosodic structure, which is appropriate in a restricted number of contexts only.

The topic-focus contour that we used in our experiment has the same contour as a broad focus one. Both have a rising pitch accent on the subject, and a falling accent on the focused word (the ‘focus exponent’). We chose a topic-focus environment instead of a broad focus one

because of the slightly clearer accent pattern produced with a topic and a focus. Even though we did not include a broad focus context in our experiment, we are confident that the pattern we call TF would get high scores in it.

4.2 Material

Three different kinds of sentences served as our experimental material: 6 short sentences, 6 long sentences and 3 sentences with ambiguous scope of negation and quantifier. Every sentence was inserted in three or four matching contexts (see below). In (7) to (9), an example for each sentence is given along with their contexts. The remaining sentences are listed in the appendix.

(7) Short sentences

Maler bringen immer Bilder mit.

Painters bring always pictures with

a. Narrow focus on the subject (NFS): Tom hat mir erzählt, dass Fotografen unserer Nachbarin immer Bilder mitbringen. Aber das stimmt nicht:

'Tom told me that photographs always bring pictures to our neighbor. But this is not true:'

b. Narrow focus on the object (NFO): Angeblich bringen Maler unserer Nachbarin immer Bücher mit. Aber das stimmt nicht:

'It is said that painters always bring books to our neighbor. But this is not true:'

c. Topic-focus (TF): Meine Nachbarin schmeißt oft große Partys, dafür bekommt sie aber auch viele Geschenke. Regisseure schenken ihr Filme, Schriftsteller Bücher und…

'My neighbor often throws big parties, and therefore she also gets lots of presents. Movie directors give her movies, writers give her books and ...'

(8) Long sentences

Passagiere nach Rom nehmen meistens den späten Flug.[4]

Passengers to Rome take mostly the late flight

a. Narrow focus on the subject (NFS): Angeblich nehmen die Leute nach Athen meistens den späten Flug. Aber das stimmt nicht:

'It is said that the people (flying) to Athens mostly take the late flight, but this is not true:'

b. Narrow focus on the object (NFO): Mona sagt, dass Passagiere nach Rom meistens die frühe Maschine nehmen. Aber das stimmt nicht:

'Mona says that passengers to Rome mostly take the early flight, but this is not true:'

c. Topic-focus (TF): Pendler, die ziemlich weit von zuhause arbeiten, haben oft ähnliche Angewohnheiten. Geschäftsleute Richtung Paris fahren oft mit dem Auto, Reisende nach London nehmen den Zug aus Calais und…

'Commuters who work far away from home often have similar habits. Business people who go to Paris often take their car, travelers to London take the train from Calais and ...'

(9) Quantifier-negation sentences:

<u>Beide Autos sind nicht beschädigt worden.</u>

Both cars were not damaged

a. Two foci ('two'): Es wäre schlimm gewesen, wenn Karl bei dem Unwetter seinen Jaguar und seinen Porsche auf einmal verloren hätte, aber glücklicherweise war es nicht so.

'It would have been too bad if Charles had lost both his Jaguar and his Porsche because of the bad weather, but fortunately this was not the case.'

b. Narrow focus on the quantifier (FQ): Ist nur Peters Auto nicht beschädigt worden? Nein,…

'Has only Peter's car not been damaged? No, ...'

c. Narrow focus on the negation (FN): Ich habe gesehen, dass Deine beiden Autos seit Wochen in der Garage stehen. Sind sie bei dem Unfall beschädigt worden? - Nein, ich habe Dir doch schon gesagt:

‘I have seen that both your cars have been sitting in the garage for ages. Were they damaged in the accident? - No, I already told you, ...’

d. Topic-focus (TF): Bei dem Unfall ist verschiedenes passiert. Drei Fahrräder sind jetzt Schrott, ein Fußgänger ist im Krankenhaus, aber bei den Autos, die dabei involviert waren, war es nicht dramatisch:

‘Several things happened at the accident. Three bikes are now ruined, a pedestrian is at the hospital, but nothing dramatic happened to the cars involved:’

Contexts and stimuli sentences were spoken by a trained speaker and recorded in a sound-proof booth on a DAT recorder. The speaker was instructed to speak naturally, in a normal tempo. He read the context-target pairs at once, first the context and then the stimulus sentences. There were 48 matching pairs for the three experiments altogether (6 short sentences, 6 long sentences and 3 quantifier-negation sentences in their contexts, thus 18 + 18 + 12 pairs). All pitch accents of a specific type were realized similarly (see figures 7.1 to 7.3 for illustrations), and controlled carefully with the help of the speech analysis program PRAAT. Several recording sessions were necessary. The sentences were evaluated by three independent trained phonologists as to their naturalness. Context sentences and stimulus sentences were digitized into individual sound files, ready to be cross-spliced. No manipulation whatever was undertaken in order to not endanger the naturalness. We prepared 36 non-matching pairs for the short sentences, 36 for the long sentences and 32 for the scope sentences = 104 non-matching pairs. The sentences to be evaluated consisted thus in 48 matching and 104 non-matching pairs = 152 pairs.

4.3 Subjects

Four non-overlapping groups of 15 subjects (altogether 60 students at the University of Potsdam) took part in the experiment. They were native speakers of Standard German and had

no known hearing or speech deficit. All were paid or acquired credit points for their participation in the experiment. Two groups judged the sentences on a scale of 1 (very bad) to 8 (perfect), and two groups judged the same sentences in a categorical way: acceptable (yes) or non-acceptable (no). All 60 informants evaluated the scope sentences. In addition, the first and third groups also judged the short sentences, while the second and fourth groups judged the long sentences, thus 30 matching sentences plus 68 non-matching ones each.

4.4 Procedure

The subjects were in a quiet room with a presentation using the DMDX experiment generator software developed by K. and J. Forster at the University of Arizona. The experimenter left the subject alone in the room after brief initial instructions as to beginning and ending the session. The subjects worked through the DMDX presentation in a self-paced manner. It lead them through a set of worded instructions, practice utterances, and finally the experiment itself, consisting of 102 target sentences. No fillers were inserted, but three practice sentences started the experiment. This experiment was itself included in a set of experiments in which the subjects performed different tasks: production of read material, and dialogues. The instructions made it clear that the aim of the experiment was to test the intonation and stress structure of the sentences, and not their meaning or syntax. The stimuli were presented auditorily only: pairs of context and stimulus sentence were presented sequentially. The subject heard first a context, and after hitting the return key, the test sentence. The task consisted in judging the adequacy of the intonation of the sentence in the given context. Every recorded sentence of the groups of short and long sentences was presented 9 times, in three different intonational and stress patterns, and each of these patterns in three different contexts. The scope sentences were presented 16 times each, in all possible variants.

The sentences were presented in a different randomized order for each subject. The set-up and the instructions included the option of repeating the context-stimulus pair for a

particular sentence. Most subjects made occasional use of this possibility. Only the last repetition was included in the calculation of the reaction time (see section 4.9).

4.5 Short and long sentences

There were six short sentences like the one illustrated in (7), consisting of a simple subject (an animate noun in plural), a verb (*mitbringen* 'bring'), an adverb (*immer* 'always') and a simple object (an inanimate noun in plural). The separable but unstressed particle *mit* was located at the end of the sentence, resulting in a non-final object. The sentences were inserted in three different contexts inducing the following information structures: narrow corrective focus on the subject (NFS), see figure 7.2 above, narrow corrective focus on the object (NFO), see figure 7.3 below, and topic-focus (TF), the unmarked prosodic structure, see figure 7.1 above. The sentences with narrow focus were elicited by replacing a pre-mentioned element with another one. Our decision to use a corrective narrow focus was driven by the intention to have a very clear accentual structure. A topic-focus was elicited by pre-mentioning some pairs of elements with the same structure as the tested sentence.

Figure 7.3 displays a narrow focus on the object. The subject *Ruderer* has a rising prenuclear pitch accent with a much smaller excursion than in the unmarked topic-focus configuration. The object carries the high-pitched nuclear accent.

INSERT FIGURE 7.3. HERE

4.6 Results and discussion

Table 7.1 displays the data for the first group of subjects, who had to give scalar judgments. Each cell shows the mean score of the six sentences having the same context-intonation pair.

The second group of subjects judged the same sentences in a categorical way, and the mean scores for these subjects are given in table 7.2. The correlation between the mean scores in table 7.1 and 7.2 is almost perfect (Pearson's product-moment correlation = 0.984, p = 0.000). The interaction between context and intonation is displayed graphically in figure 7.4. It presents the results of only the first group (i.e. scale answers), but a graph of the second group would look very similar due to the strong association between the two groups.

INSERT TABLE 7.1 HERE

INSERT TABLE 7.2 HERE

INSERT FIGURE 7.4 HERE

All patterns were accepted best in their own matching context. The unmarked TF tonal contour, corresponding to the UPS, was also readily accepted in the NFO context, a result corresponding to our expectations. NFO had one pitch accent on the object and a reduced prenuclear accent on the subject. It thus looked more like the TF (the realization of the UPS) than the NFS with only one pitch accent on the subject. NFO got intermediate scores in the TF context. The slight inadequacy that our informants felt can be safely attributed to the lack of a topical accent on the subject. By contrast, NFS is accepted in its matching context, but refused in a non-matching context.

Gradient judgments were obtained in two different ways, either directly, by letting the informants give their own gradient results, or indirectly, by counting categorical results. The very high correlation between the two groups of means suggests that it does not matter which method is used, as both methods give very similar results. It will be shown that this correlation reproduced itself for all sentences.

In the six longer sentences, one of which is illustrated in (8), the subject and the object were syntactically more complex. We decided to include both short and long sentences in our experiment in order to verify the influence of length and complexity on the perception of tonal patterns. The distinction between the two kinds of sentences, however, turned out to be minimal, as one can see from a comparison between figure 7.4 and figure 7.5.

INSERT FIGURE 7.5 HERE

The only difference between these sentences and the short ones worth mentioning is that in the TF context, both NFS and NFO were now better tolerated. We do not have any explanation for the slightly better acceptance of the absence of a late accent in a TF context. As an explanation for the better acceptance of NFO in the TF context, we offer that it might be not so easy to perceive the difference between weak and strong prenuclear accents when the sentence is longer. In the absence of concrete hypotheses and more stable results, we refrain from speculating about the reasons for this difference.

Here also a very high correlation between the two groups of subjects was found, suggesting once more that both scalar and categorical methods are equally good for obtaining gradient judgments.

Let us now relate our findings to those described in section 2. First, the scores for matching context-intonation pairs were higher than for non-matching pairs. Second, a missing nuclear accent and an added nuclear accent triggered lower scores than sentences with the expected accentuation. The same was true for both a missing prenuclear accent and an added prenuclear accent. As described by Hruska et al. (2001), adding a prenuclear accent on the subject in a situation where only a nuclear accent on the object is expected obtained higher scores than other non-matching pairs. In the same way, Gussenhoven, as well as Clifton & Birch, also found that an added prenuclear accent delivers better judgments than an added nuclear accent.

4.7 Scope sentences

The sentences in the third experiment, one of which is illustrated in (9), consist of a subject made up of a quantifier and a noun, an auxiliary, the negation *nicht*, and a past participle or an adjective (below called 'the predicate'), and are characterized by variable scope of negation and variable scope of the quantifier. Four contexts were constructed, as illustrated in (9). First a context eliciting two accents: one on the quantifier and one on the negation (called 'two' in the following). The second context elicits a narrow focus on the quantifier (FQ), the third context a narrow focus on the negation (FN), and the last context was a topic-focus one, eliciting two accents again, one on the quantifier, as in 'two', and the second one on the predicate (TF). All four contours are illustrated for example (9) in figure 7.6.

INSERT FIGURE 7.6 HERE

The syntactic structure of the sentences in this experiment is simple, but their semantic structure is not. First, the negation can have scope over the quantifier or, vice-versa, the

quantifier can have scope over the negation. In the experiment, one context called unambiguously for wide scope of the negation ('not both cars…'), and one unambiguously for wide scope of the quantifier ('for both cars, it is not the case that…'). The first case ('two' context in 9) is triggered by double accentuation on the quantifier and the negation, and the second case (FQ context in 9) comes with a single accent on the quantifier.[5] It is assumed here that the scope inversion reading elicited by the 'two' context can be explained by general properties of topicalization, visible in languages with resumptive pronouns. The topicalized quantifier in the sentences under consideration is in a position of extraposition to the left, but is nevertheless interpreted to be in the scope of the negation (see also Höhle 1991). All authors who have studied the scope inversion phenomenon in German (Höhle 1991, Jacobs 1997, Büring 1997, Krifka 1998) have insisted on the necessity of a rise-fall contour to get the interpretation aimed at, and this is the contour which was produced by our speaker as well. Crucially, an independent phonological phrase is formed which contains the topicalized constituent, separate from the main clause. In a realization with only one accent on the quantifier, by contrast, both the quantifier and the negation are interpreted in situ and consequently, the quantifier has wide scope over the negation.[6] Prosodically, the quantifier cannot be interpreted as being topicalized because it has the focal accent of the sentence. In our experiment, the context eliciting this accent pattern was one in which the quantifier was contrastively accented.

The other two patterns, a single accent on the negation (FN) and a double accent on the quantifier and on the predicate (TF) do not evoke clear scopal relationships. A unique accent on the negation contradicts the preceding sentence. In the experimental sentences, the predicate had been stressed in the preceding matching context. However, it was not possible to unambiguously reconstruct the context from the negated sentence only. An accent on the quantifier, the noun or the predicate changes the pragmatics of the sentence, but in the

realization with a single accent on the negation, these differences are cancelled. The hypothesis was thus that an accent on the negation would be tolerated in a variety of contexts.

The TF context with accents both on the NP containing the quantifier and on the predicate is similar to the 'two' context. It can also have different readings, one being that the predicate is contrasted. Inverted scope is also not impossible in this case.

To sum up, a realization with a single accent—especially when the accent is on the quantifier—seems to be more marked than a realization with two accents, in the sense that it is adequate in less contexts. With the third experiment, we wanted to verify this hypothesis.

4.8 Results and discussion

Tables 7.3 and 7.4 as well as Figure 7.7 present the mean values in both scalar and categorical judgments. Once again, the correlation between the two groups of means is almost perfect (Pearson's product-moment correlation = 0.973, p = 0.000).

INSERT TABLE 7.3 HERE

INSERT TABLE 7.4 HERE

INSERT FIGURE 7.7 HERE

The results are not as clear cut as in the short and long sentences. For the 'two' context, the FN and the FQ, the matching pairs obtained better scores than the other ones. It is also to be noticed that the TF and 'two' contexts are nearly interchangeable. This can be

attributed to the presence of two accents in both sentences, fitting both contexts requiring two accents. The same cannot be said for the realizations with one accent since the accent elicited in each case is at a different place. However, the FN sentences, with a late accent, elicited better scores in non-matching environment than the FQ sentences with an early accent. The highly marked prosodic pattern found in FQ sentences obtained poor scores in all non-matching contexts, and the best results in the matching context.

To sum up the results obtained for these sentences, it can be observed that the interchangeability of contexts and intonation pattern is higher in these sentences than in the short and long sentences. We explain this pattern of acceptability with the fact that the scope structure of these sentences, complex and subject to different interpretations, renders the accent patterns less rigid. Another interpretation could be that speakers were more concentrated on understanding the scopal relationships and were thus less sensitive to slight variations in the tonal structure of the sentences they heard.

4.9 Reaction times

Additional information on the cognitive cost of the task was gathered by the measure of reaction times. Table 7.5 shows that it took more time to process the long sentences and the scope sentences than the short ones. It can also be observed that making a decision on a scale needs more time than making a categorical decision (except for the long sentences, where no difference could be observed). We could not find any correlation between the number of keys available for responding and the reaction times, neither in the scalar decision task when comparing the subjects who used all keys and those using only four to six keys (out of the eight at their disposal), nor between the two tasks in comparison. In other words, it is not the case that using eight keys instead of two increases the time it takes to make a decision. We conclude that the increase of reaction time that we observe is truly due to an increase of cognitive complexity.

INSERT TABLE 7.5 HERE

---

5 Conclusion

This chapter has investigated the gradient nature of the acceptability of intonation patterns in German declarative sentences. Three kinds of sentences elicited in different information structural contexts were cross-spliced and informants were asked to judge the acceptability of context-target pairs. The clearest results were obtained for the short sentences, though the long sentences delivered comparable results. Finally, the tonal patterns of scope sentences were much more difficult to interpret, because the scope behavior of the negation and the quantifier was variable, depending on the accent structure of these sentences. For all sentences, we found that a prosody with two accents got better scores than a prosody with only one accent, and that a contour with a late accent was better accepted in non-matching environments. We dubbed the prosody with two accents, acceptable in a broad focus context or in a topic-focus context, UPS, for 'unmarked prosodic structure,' and we observe that this contour is accepted in a non-matching context more readily than contours with only one accent, especially when this single accent is located early in the sentence.

The results of the short and long sentences, and, to a lesser extent, those of the scope sentences, point to a good correlation between context and prosodic structure. Speakers and hearers do use prosodic information such as presence vs. absence of pitch accents, their form and the phrasing to assess the well-formedness of context-target sentence pairs, and they do so consistently. Their performance is ameliorated when the syntactic and semantic structure of the sentence is very simple. It can safely be claimed that in German, information structure

plays an important role in the processing of prosody, whereas it has been shown for syntax that word order alone, presented in written form, does not have the same effect (see for instance Schlesewsky, Bornkessel & McElree, this volume, and references cited there). The conclusion one could tentatively draw from this difference is that intonation encodes information structure better than syntax.

An interesting result is that in all three experiments the scores obtained for the two groups of subjects (scale and yes-no answers) were similar. In other words, the same gradient results can be obtained by using either gradient or not-gradient judgments. This is remarkable since the cognitive task executed in both groups was different. It could have been the case that in a sentence with a high score of acceptability the rating by scale would have been gradient, but the yes-no judgment categorical. However, if the groups of informants are large enough, 'intolerant' subjects compensate for the degree of insecurity that remains in subjects asked to give a judgment on a scale.

Though we offer no analysis of how our gradient data can be accounted for in a formal grammar, we conclude with the observation that a categorical grammar will not be adequate. Speakers are more or less confident in their judgments, and gradiently accept sentences intended to express a different information structure, depending on whether the sentences have a similar accent pattern. A gradient grammar, like stochastic OT, which uses overlapping constraints, can account much better for the observed variability. This is, however, a subject for future research.

Appendix

Short sentences (three contexts)

1. *Maler bringen immer Bilder mit*. 'Painters always bring pictures.'

2. *Lehrer bringen immer Hefte mit*. 'Teachers always bring notebooks.'

3. *Sänger bringen immer Trommeln mit*. 'Singers always bring drums.'

4. *Ruderer bringen immer Boote mit*. 'Oarsmen always bring boats.'

5. *Geiger bringen immer Platten mit*. 'Violinists always bring records.'

6. *Schüler bringen immer Stifte mit*. 'Students always bring pens.'

Long sentences (three contexts)

7. *Passagiere nach Rom nehmen meistens den späten Flug*.

'Passengers to Rome always take the late flight.'

8. *Reisende nach Mailand fahren oft mit dem schnellen Bus*.

'Travelers to Milan often travel with the express bus.'

9. *Autofahrer nach Griechenland nehmen immer den kürzesten Weg*.

'Car drivers always take the shortest road.'

10. *Schiffe nach Sardinien fahren meistens mit voller Ladung*.

'Ships to Sardinia mostly sail with a full cargo.'

11. *Züge nach England fahren oft mit rasantem Tempo.*

'Trains to England often ride at full speed.'

12. *Trekker nach Katmandu reisen meistens mit vollem Rucksack.*

'Trekkers to Katmandou mostly travel with a full backpack.'

Variable scope sentences (four contexts)

13. *Alle Generäle sind nicht loyal*. 'All generals are not loyal.'

14. *Beide Autos sind nicht beschädigt worden*. ‘Both cars have not been damaged.’

15. *Viele Gäste sind nicht gekommen*. ‘Many guests did not come.’

Figure 7.1: **Pitch track of *RUDERER bringen immer BOOTE mit***

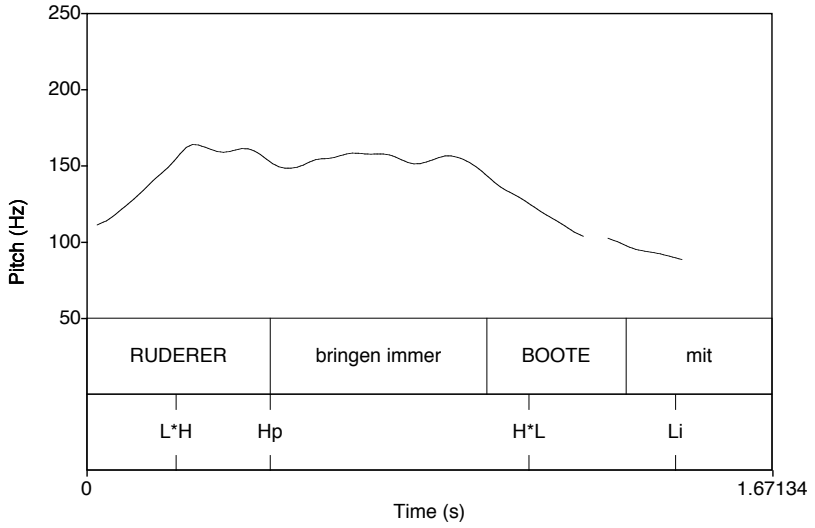


Figure 7.2: **Pitch track of** *RUDERER bringen immer Boote mit*

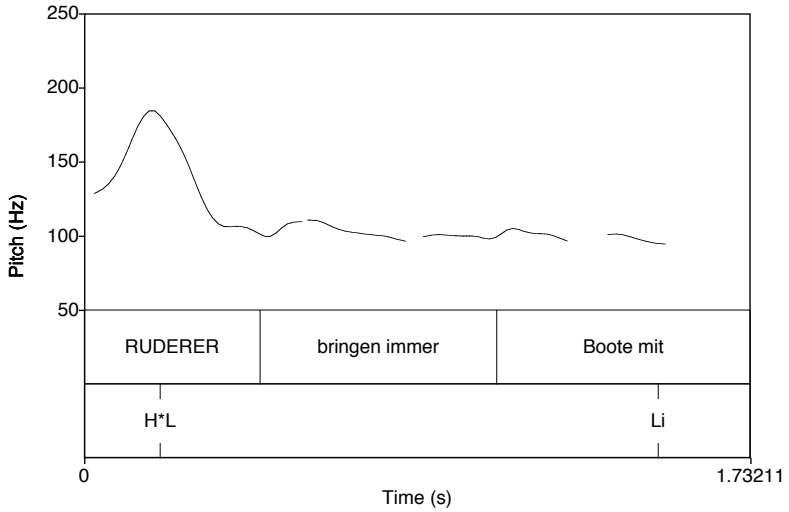


Figure 7.3: **Pitch track of** *Ruderer bringen immer Boote mit*

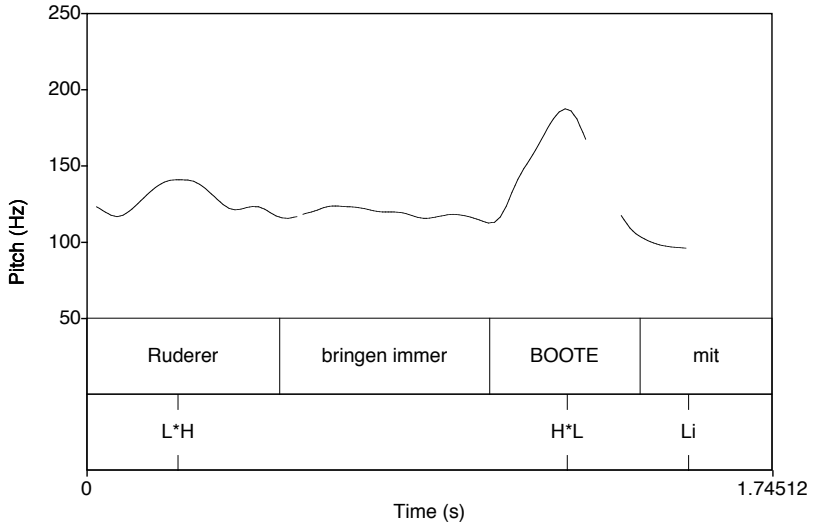


Figure 7.4: **Mean acceptability scores for short sentences** (scale answers)

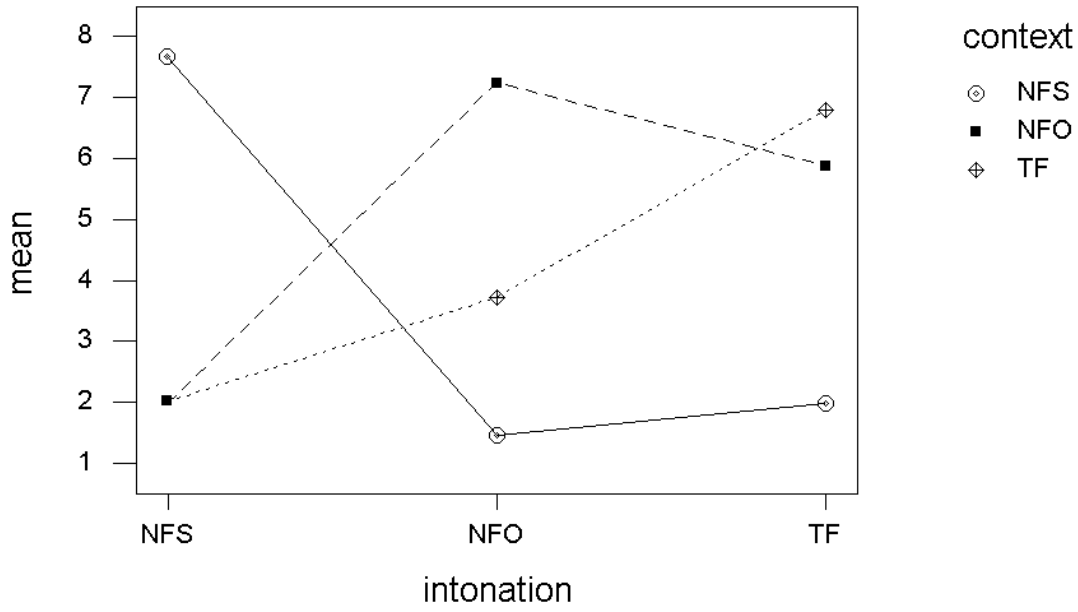


Figure 7.5: **Mean acceptability scores for long sentences** (scale answers)

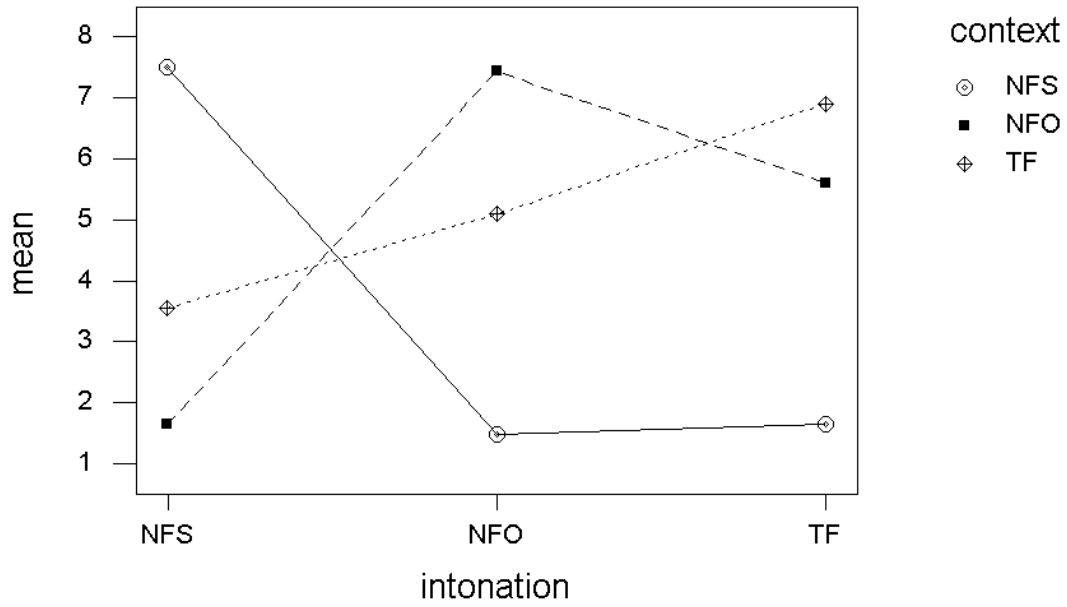
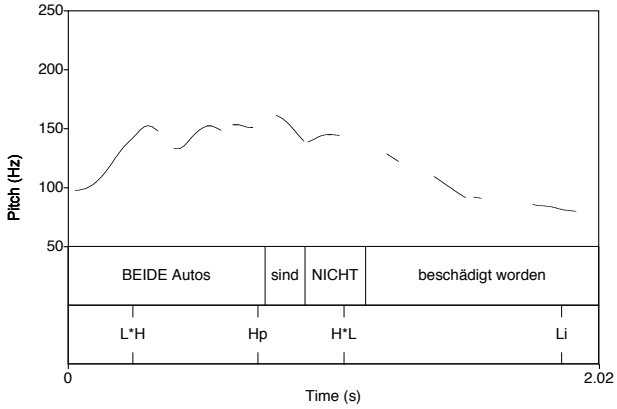
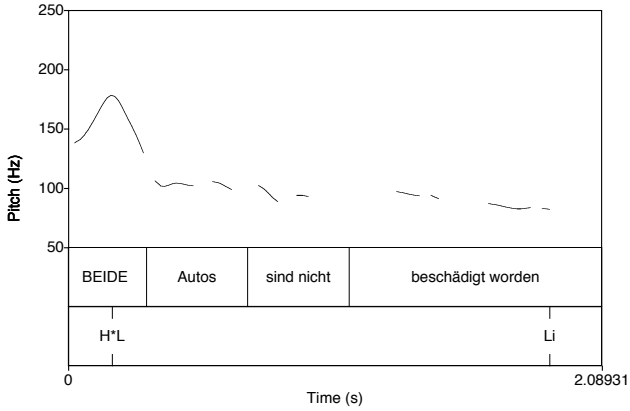


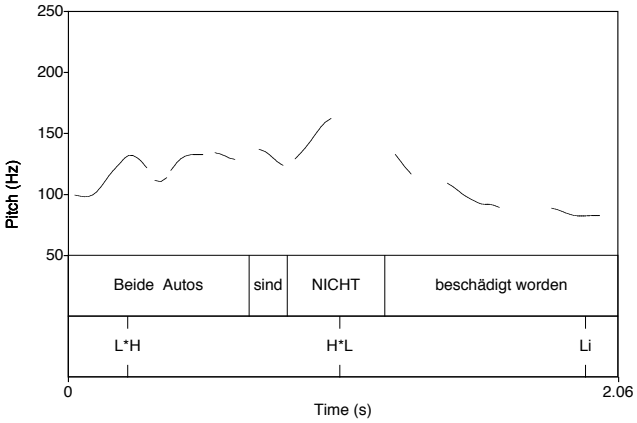
Figure 7.6: **Four realizations of** (9)



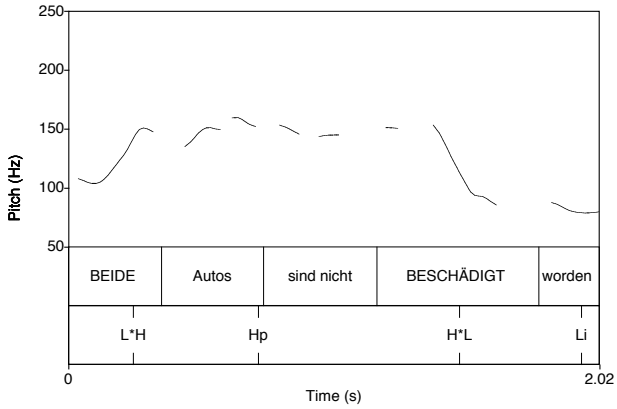
i. Context 'Two'



ii. Context 'FQ'



iii. Context 'FN'



iv. Context 'TF'

Figure 7.7: **Mean acceptability scores for scope sentences** (scale answers)

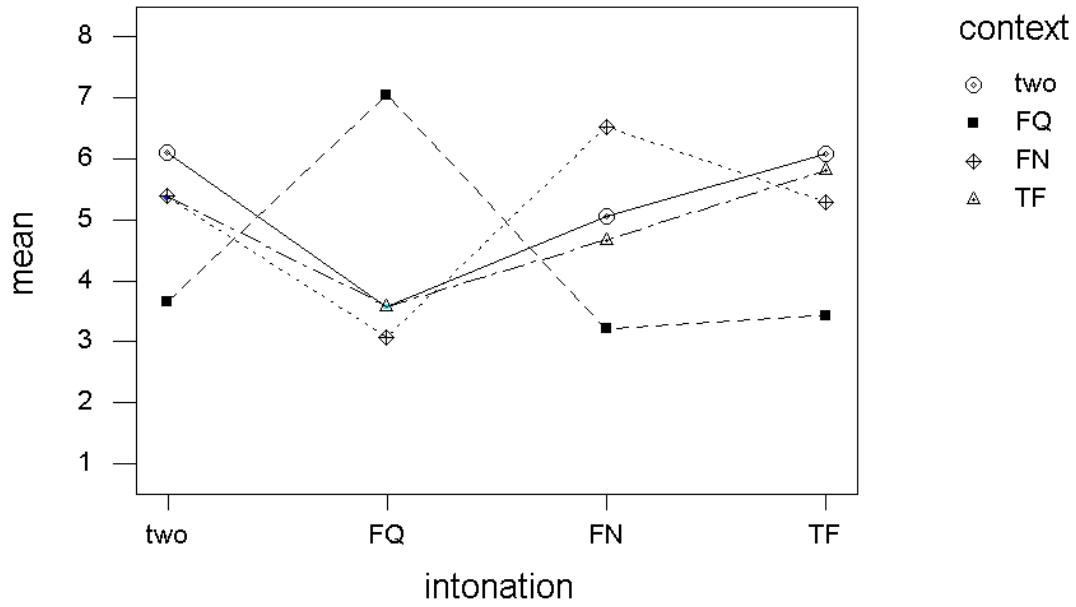


Table 7.1: **Short sentences: mean judgment scores** (on a scale from 1 to 8)

Context / intonation	NFS	NFO	TF
NFS	7.7	1.5	2.0
NFO	2.0	7.2	5.9
TF	2.0	3.7	6.8
All contexts	3.9	4.1	4.9

Table 7.2: **Short sentences: mean judgment scores** (categorical).

Context / intonation	NFS	NFO	TF
NFS	0.92	0.18	0.11
NFO	0.22	0.89	0.66
TF	0.07	0.32	0.87
All contexts	0.40	0.46	0.54

Table 7.3: **Scope sentences: mean judgment scores** (on a scale from 1 to 8)

Context / intonation	two	FQ	FN	TF
two	6.1	3.6	5.1	6.1
FQ	3.7	7.0	3.2	3.4
FN	5.4	3.1	6.5	5.3
TF	5.4	3.6	4.7	5.8
All contexts	5.1	4.3	4.9	5.1

Table 7.4: **Scope sentences: mean judgment scores** (categorical)

Context / intonation	two	FQ	FN	TF
two	0.73	0.27	0.64	0.71
FQ	0.32	0.90	0.26	0.36
FN	0.62	0.18	0.90	0.57
TF	0.76	0.39	0.54	0.72
All contexts	0.61	0.43	0.59	0.59

Table 7.5: **Mean reaction times**

	Short sentences	Long sentences	Scope sentences
Scale	4.2 s	5.0 s	5.0 s
	(sd = 2.24; N = 810)	(sd = 2.36; N = 810)	(sd = 2.83; N = 1440)
Categorical	3.7 s	5.0 s	4.7 s
	(sd = 2.22; N = 810)	(sd = 2.27; N = 810)	(sd = 2.80; N = 1440)

[1] A pilot experiment for this paper was presented at the Potsdam Gradience Conference in October 2002 and some of the results discussed here were presented at the Syntax and Beyond Workshop in Leipzig in August 2003. Thanks are due to two anonymous reviewers, as well as to Gisbert Fanselow and Ede Zimmermann for helpful comments. Thanks are also due to Frank Kügler for speaking the experimental sentences, and to Daniela Berger, Laura Herbst and Anja Mietz for technical support. Nobody except for the authors can be held responsible for shortcomings.

[2] Some have found categories in the domain of pitch accent realization, like for instance Pierrehumbert & Steele (1987) or Ladd & Morton (1987).

[3] Birch & Clifton's results also indicate that a single accent on the verb is readily accepted in a context eliciting broad focus (78% of yes). The only situation where speakers accepted a pair less (with 54%, and not between 71 and 84% as in the other pairs) was when the context was eliciting a narrow focus on the verb, and the answer had a single accent on the argument (QB/R2).

[4] As Ede Zimmermann (p.c.) observes, it is not undisputed whether there is a structural ambiguity between the temporal and the quantificational reading of *meistens*. We suspect that, even if confirmed, this ambiguity played no role in the experimental results.

[5] As a generalization, the negation may have wider scope when both the quantifier and the negation (or the negated constituent) are accented. This generalization holds only for this type of constructions, but not for other sentences with inverted scope, like those with two quantifiers discussed in Krifka (1998).

[6] Krifka (1998) explains scope inversion of sentences with two quantifiers by allowing movement of accented constituents at the syntactic component of the grammar. Both

topicalized and focused constituents have to be pre-verbal at some stages of the derivation in order to get stress.